

Kredi Kartı Dolandırıcılık Tespiti

- ▶ FET445 VERİ MADENCİLİĞİ PROJESİ
 - ▶ GRUP: DETECTIONGROUP
 - ▶ MUHAMMET AMROUŞ (22040301192)
MOHAMED KHALED HAJASAAD (22040301156)
 - ▶ RAKAN HEJAZI (222040101119)
 - ▶ KASIM ŞEHUNI (22040301228)
 - ▶ MOHAMAD NOUR NASIF (22040301095)
- VİDEO LİNKİ:
- <https://www.youtube.com/watch?v=1K0dNWKzkMY>

Problemin Açıklaması

- Problem: Finansal işlemlerdeki gürültülü veriler arasında sahteciliği anlık tespit etmek.

- Temel Zorluk (Class Imbalance): Veri setinde dolandırıcılık oranı sadece %0.17'dir.

- Amaç: Standart modellerin sürekli 'normal' diyerek yanılıltıcı yüksek doğruluk vermesini engellemek, F1-Score ve Recall odaklı bir sistem geliştirmek.

Veri Seti Açıklaması

- Veri Kaynağı: Kaggle - Credit Card Fraud Detection Dataset

- Boyut: 284,807 İşlem, 31 Özellik.

- Özellikler: PCA ile dönüştürülmüş V1-V28 bileşenleri, Time ve Amount.

- Sınıf Dağılımı: 284,315 Normal (%99.83) vs 492 Fraud (%0.17).

Yöntem ve Feature Engineering

- Feature Scaling: 'StandardScaler' kullanılarak tüm değişkenler normalize edildi.

- Train-Test Split: Veri seti %80 Eğitim, %20 Test olarak ayrıldı.

- Performans Metrikleri: Accuracy (Doğruluk), Precision (Kesinlik), Recall (Duyarlılık), F1-Score ve ROC-AUC.

Best Model: XGBoost

- Model Yaklaşımı: Gradyan artırma tabanlı topluluk öğrenmesi (XGBoost).

- Hiperparametreler: 'RandomizedSearchCV' kullanılarak learning_rate, n_estimators ve max_depth tune edildi.

- Önemli Özellikler (Feature Set): V14, V10, V4 ve V12 en ayırt edici özellikler olarak belirlendi.

Modellerin Karşılaştırılması

Model	Accuracy	F1-Score	AUC
XGBoost (Best)	%99.95	0.86	0.98
Random Forest	%99.94	0.85	0.97
LightGBM	%99.93	0.84	0.97
MLP (Deep Learning)	%99.91	0.81	0.96
Logistic Regression	%99.85	0.72	0.91

Sonuç ve Değerlendirme

